The Grid for Engineers

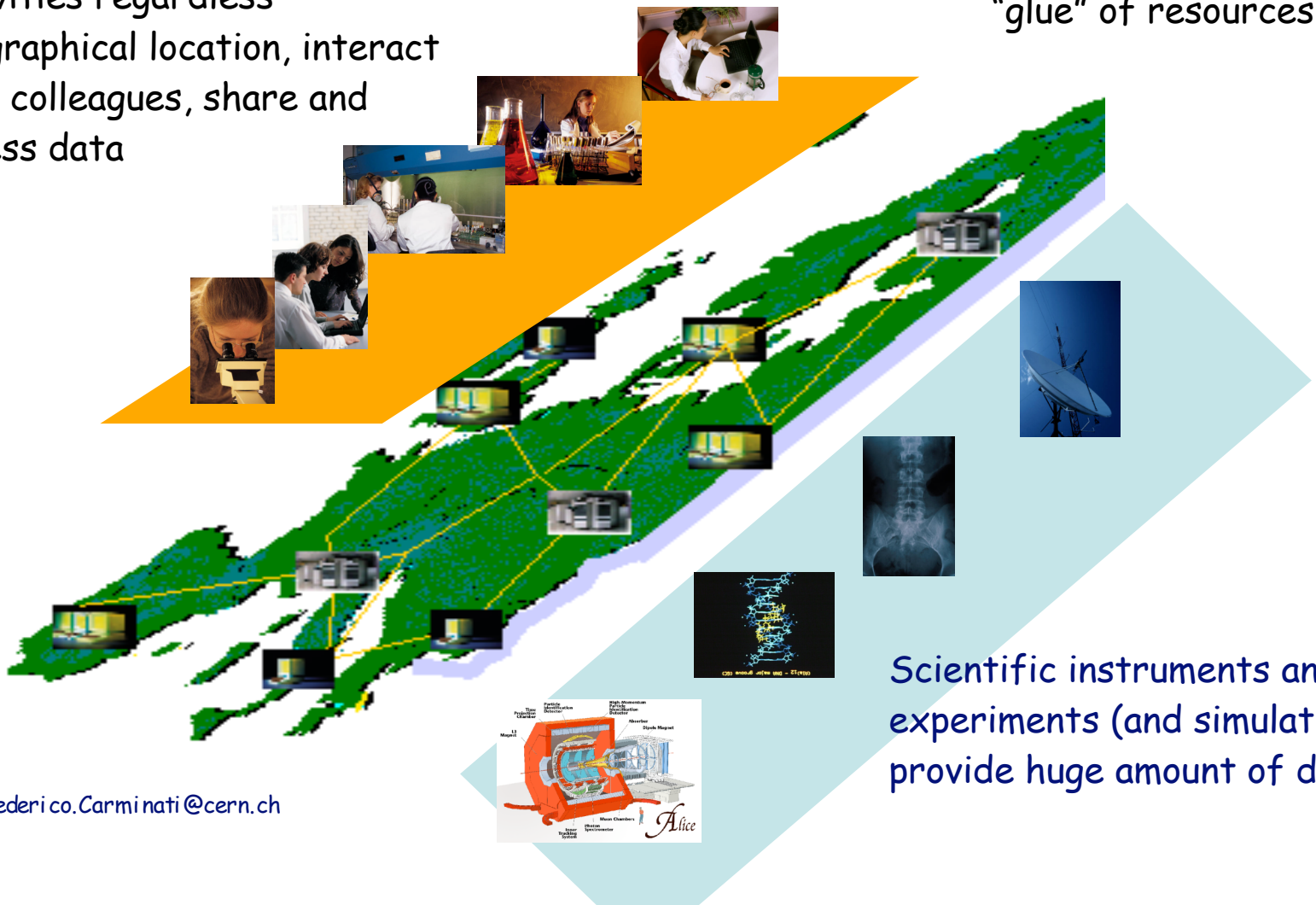
Ruth Pordes
Fermilab

With thanks for slides from Ian Foster, Vicky White, and many others

The Grid Vision

The GRID: networked data processing centres and "middleware" software as the "glue" of resources.

Researchers perform their activities regardless geographical location, interact with colleagues, share and access data



Scientific instruments and experiments (and simulations) provide huge amount of data

Federico.Carminati@cern.ch

Coordinated Sharing of Heterogenous

Computation, Storage, Network…

Resources across a set Administrative Domains by

Multiple Dynamic Organizations

Interview: The Future in Grid Computing

By David Worthington, BetaNews

February 21, 2005, 11:41 AM

INTERVIEW Computing grids are software engines that pool together and manage resources from isolated systems to form a new type of low-cost supercomputer. In spite of their usefulness, grids remained the plaything of researchers for many years. But now, in 2005, grids have finally come of age and are becoming increasingly commercialized.

Sun Microsystems recently unveiled a new grid computing offering that promises to make purchasing computer time over a network as easy as buying electricity and water. Even Microsoft is said to be investing in grids and Sony has grid-enabled its PlayStation 3 for movie-like graphics.

As interest in these distributed technologies grow, so does the probability for disinformation. With that in mind, BetaNews sat down with some of the world's leading grid guru's, Dr. Ian Foster and Steve Tuecke, to set the record straight and divorce grid hype from grid reality.

BetaNews: Since we last spoke in 2001, what significant developments have there been in the commercialization of grid technologies?

Dr. Ian Foster: Back then we were just seeing earlier interest in grid technologies from companies like IBM etc. Since then we have seen tremendous growth and enthusiasm. And a lot of things are being labeled as grid that perhaps one could argue they are not. Perhaps they are more, in some cases, computing cluster management solutions, but also some substantial early deployments in the industry from companies like IBM and Sun, and others like HP and so forth.

Then more recently we have seen Univa being created, which I am involved as founder and advisor.

Reality can Still be Pretty Simplistic - even in the Commercial World

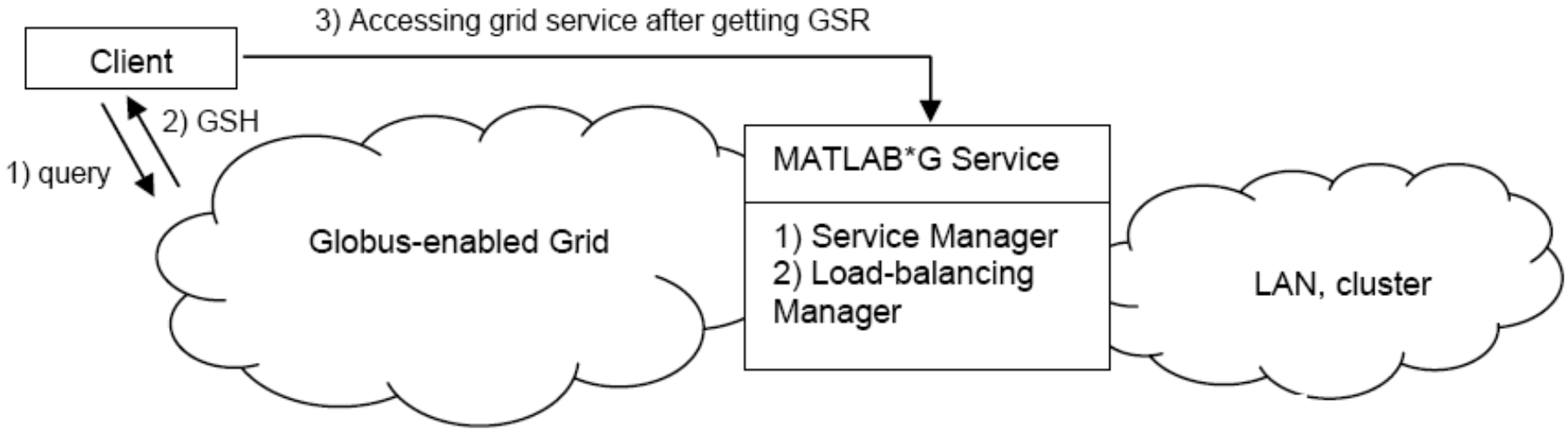


Figure 1: Accessing a Grid Service

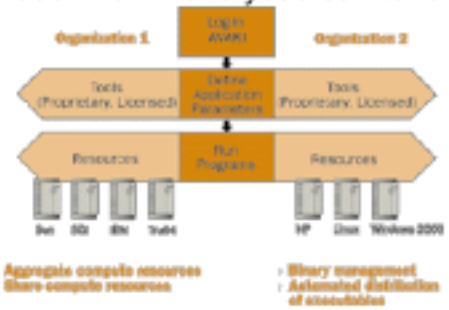
Grid Computing: More Power to You

An emerging technique for distributed computing promises better-engineered designs.

Alan Joch

The next time you see one of McLaren International's sleek Formula One racing cars nose its way to a checkered flag, think grid. For the past year, designers at McLaren's U.K. headquarters have been using emerging grid computing technology to link dozens of Sun E420R four-processor workstations into a single, virtual supercomputer. With that amount of processing muscle, McLaren's engineers can efficiently move CAD data into airflow models to perform highly complex computational fluid dynamics (CFD) analyses on new car designs.

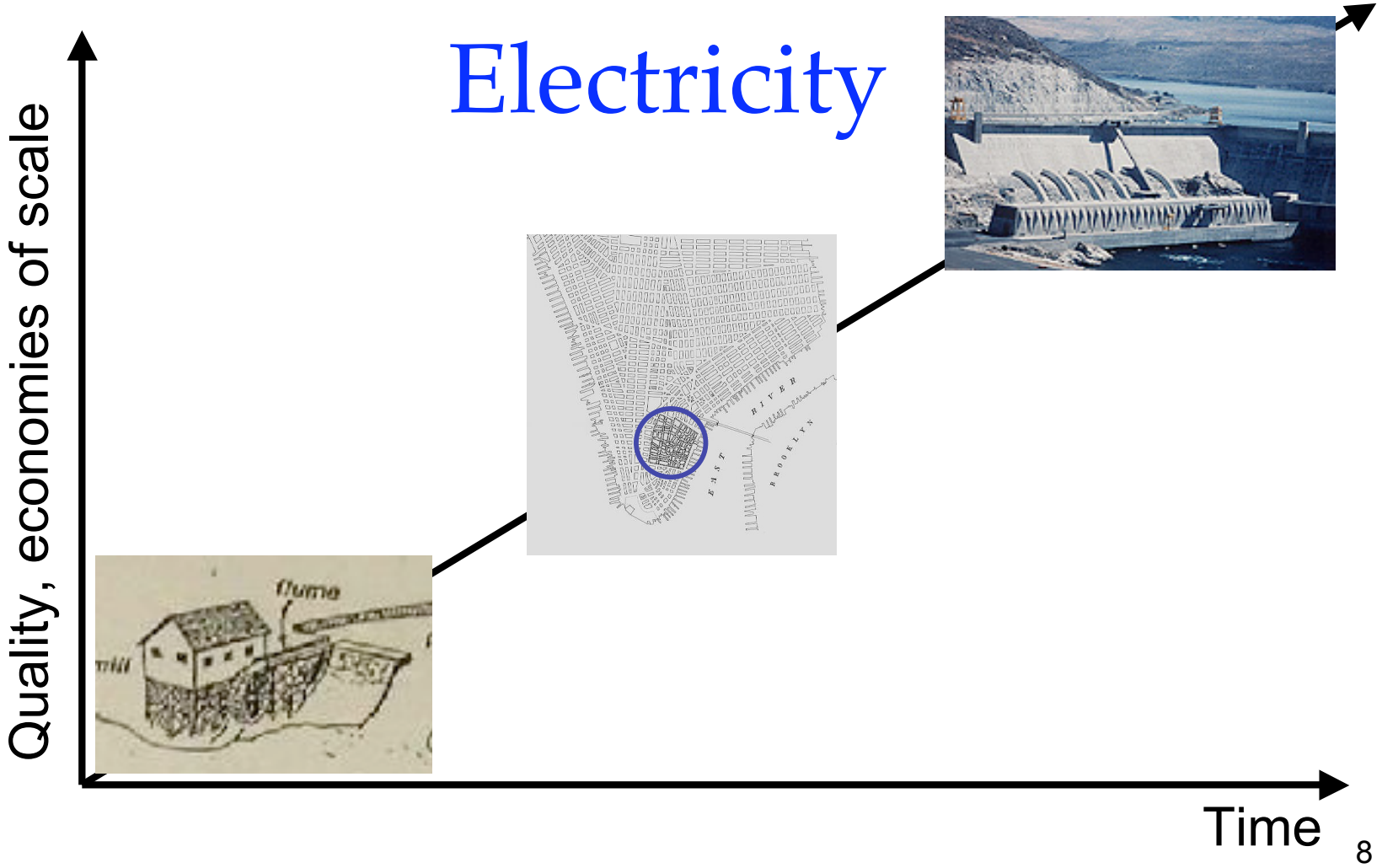
These models show the aerodynamic qualities of new body and wing shapes and help McLaren engineers devise alternative ways to make the cars slice through the air with ultimate efficiency. Because the interplay between shapes, speed, and angles creates so many data points—as many as 10 million elements in one model—CFD analyses can take weeks to perform even with high-end stand-alone computers. When McLaren's 104 processors break down an analysis into workable chunks and crunch the numbers 24/7, the results come fairly quickly—from a few hours to a couple days, depending on the job.



But reduced processing time is only a side benefit. The real value of distributing large compute jobs over a bank of computers is better-engineered designs. "When we get faster turnaround times, we can increase our design

Grid, grid, grids

The (Power) Grid: On-Demand Access to Electricity



By Analogy a Computing grid

- Decouple production and consumption
  - Enable on-demand access
  - Achieve economies of scale
  - Enhance consumer flexibility
  - Enable new devices
- Standardization of interfaces
  - Voltage, current, frequency, plugs
- On a variety of scales
  - Department, Campus, Enterprise, Internet

State of the art today for "The Grid" is

Moving data files between institutions, across continents, (almost) seamlessly and automatically

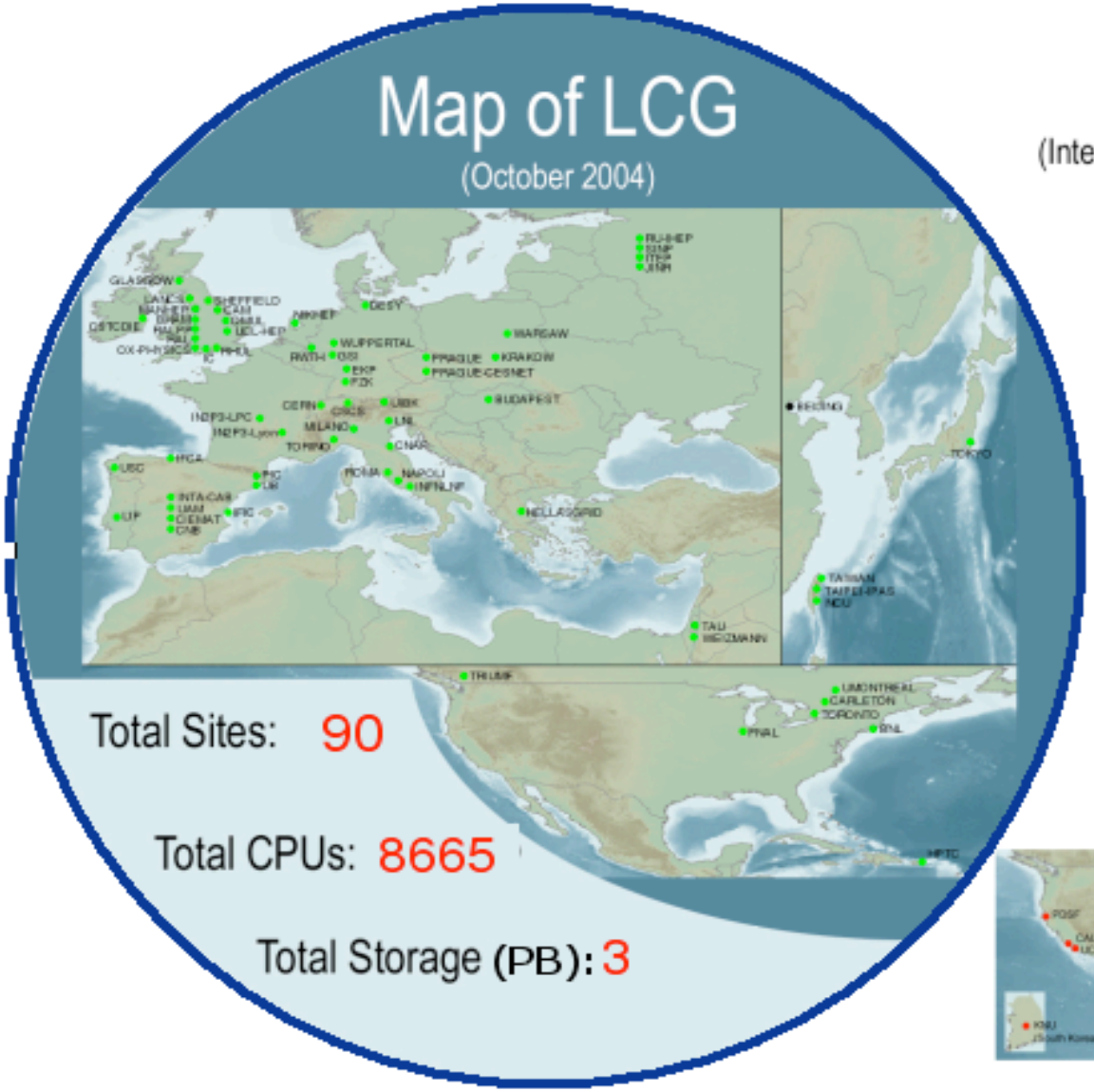
Submitting jobs from your desktop and having them run somewhere at one of the "centers" of your "virtual organization"

There are Many Grids

National

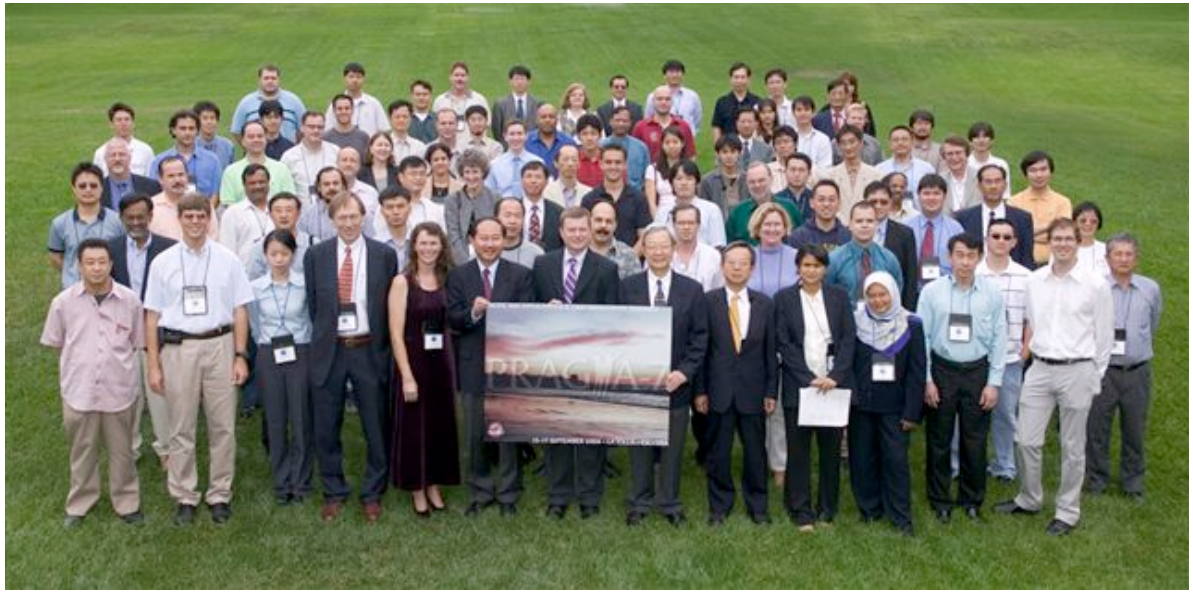
Cooperating

Federated



NorduGrid
(Interoperating with LCG)

Grid3
(Interoperating with LCG)

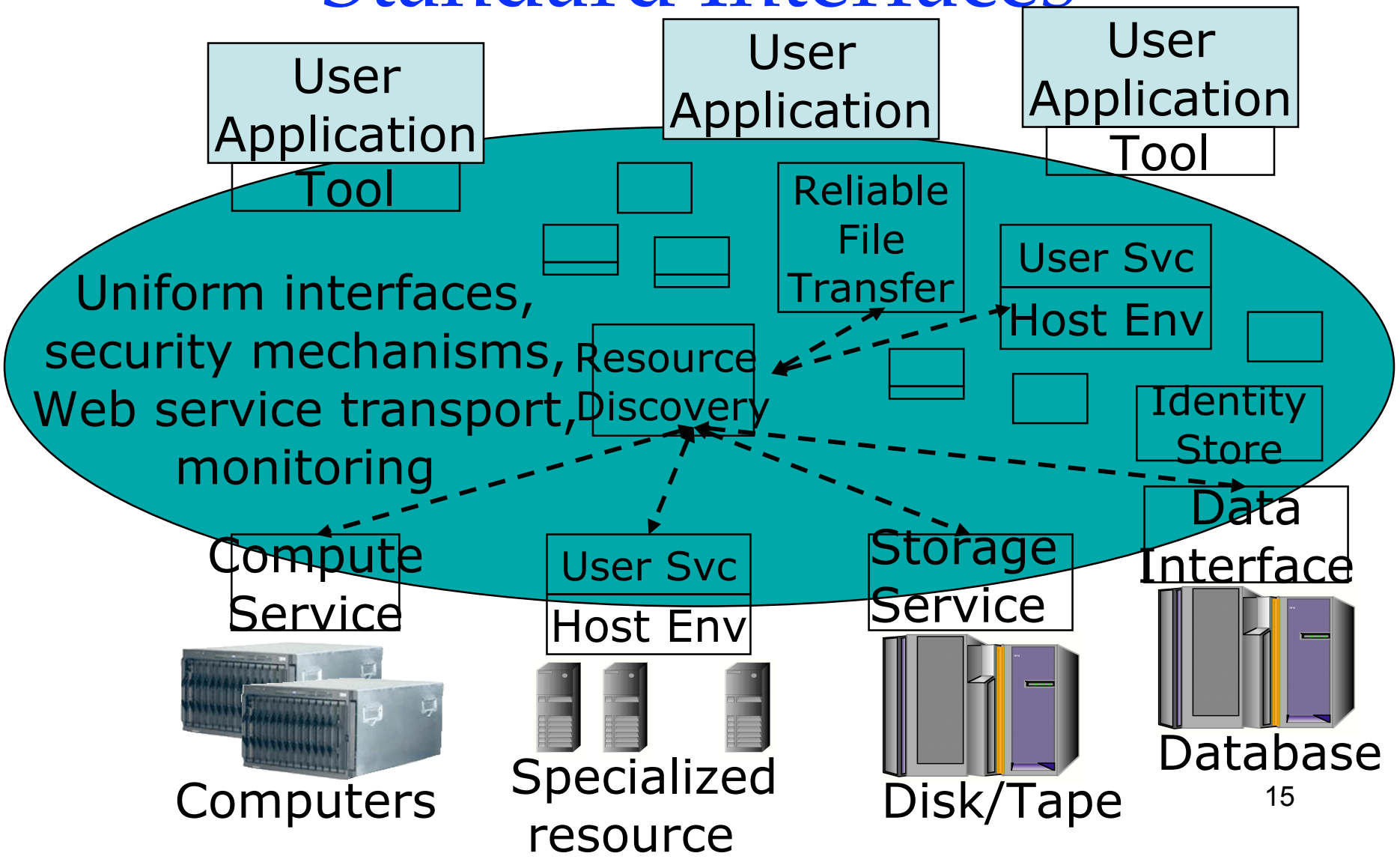


- **PRAGMA 7: Practicing Innovative, International, and Interdisciplinary Science**
- La Jolla, Calif., U.S. —The seventh annual Pacific Rim and Grid Middleware Assembly (PRAGMA) conference celebrated the achievements of its middleware projects and the integration of new testbed resources at 10 institutions, two important steps towards PRAGMA's chief goal of constructing a viable grid.
- Held on the UCSD campus from Sept. 15 through 17, PRAGMA 7 unveiled data gathered from continuously running a sample computational chemistry application over three months on 10 compute platforms in the PRAGMA testbed. "By allowing these applications to drive the underlying middleware deployment and configuration, we are learning how to share resources across international boundaries." said Mason Katz, co-chair of the Resources Working Group and co-chair of the PRAGMA 7 Workshop. "In addition, the lessons learned from this experience will help shape the construction of international production grids and will be valuable for hosts of applications."
- PRAGMA 7 also highlighted the successful integration of the Grid Datafarm distributed file system (gfarm) and the Genome Annotation Pipeline (iGAP), a suite of bioinformatics software. The project not only achieves software interoperability and provides access to more users, it also illustrates the value of researchers working across disciplines and continents, which is "critical to building a community of researchers, colleagues, friends, and ultimately an extended global family," said Dr. Jysoo Lee, deputy chair, PRAGMA Steering Committee and director of the Korea Institute for Science and Technology Information Supercomputing Center.

grid infrastructure enables

- Utility computing
- Virtualization
- Data center automation
- Adaptive enterprise
- Collaboratories

…to access resources through Standard Interfaces



Standards are needed to allow heterogeneous implementations to participate in a common infrastructure.

Can the grid standard interfaces and services be regarded as adding intelligence to the Network layer?

- Dispatch work anywhere - full connectivity)
- Discover spare compute cycles - learn about host capabilities
- Name Servers to identify grid services.

e.g. Security

Person has single "Login" and Identity Certificate (Credit Card) e.g.:

Issued to:
Subject: CN=Ruth Pordes 101995, OU=People, DC=doegrids, DC=org

Serial Number: 0E:EB

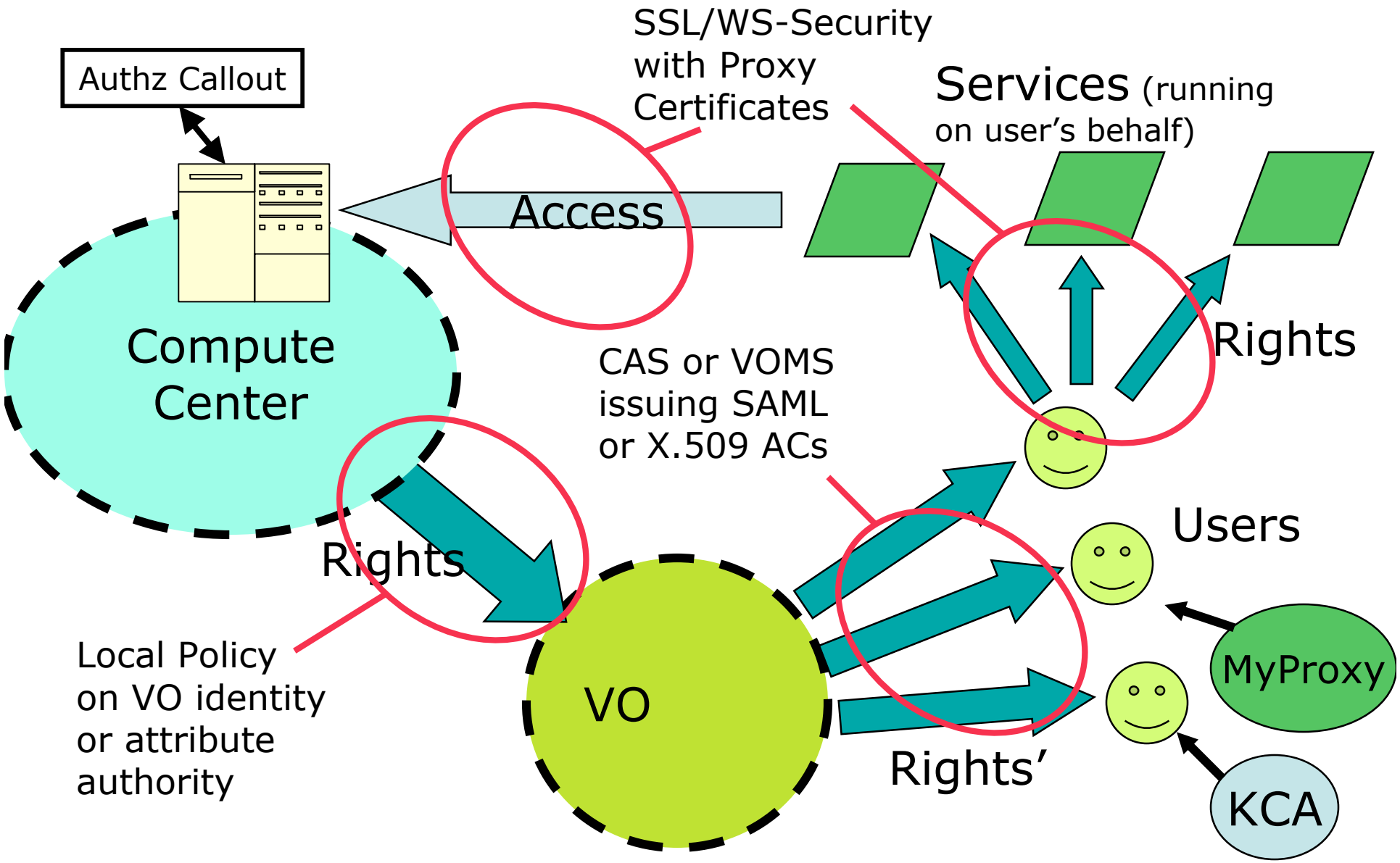
Valid from 1/26/05 8:28 AM to 1/26/06 8:28 AM

Purposes: Client,Server,Sign,Encrypt

Issued by:
Subject: CN=DOEGrids CA 1, OU=Certificate

Authorities, DC=DOEGrids, DC=org

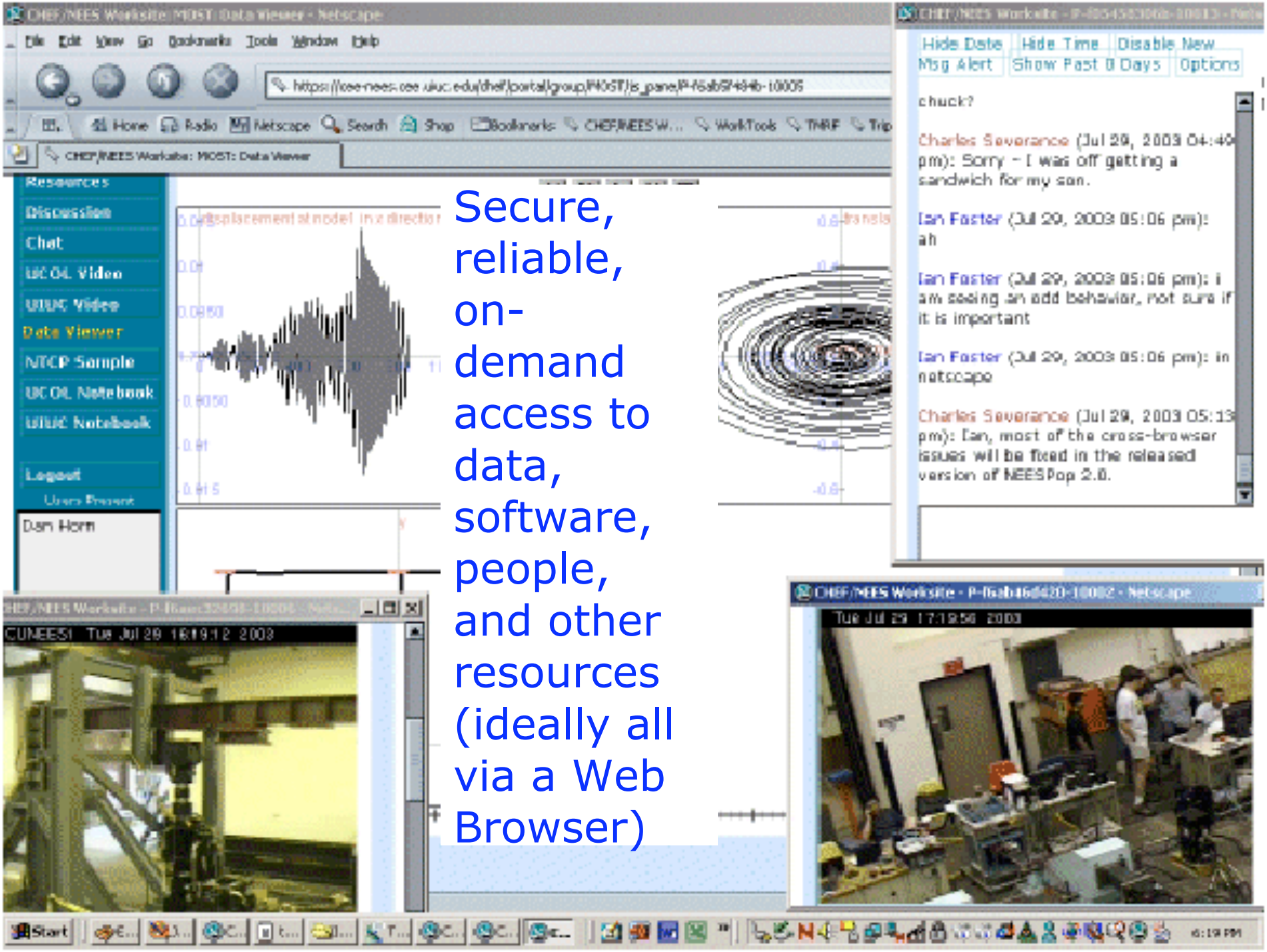
Security Details



a domain example:

Earthquake Engineering Simulation

NEESgrid Multisite Online Simulations


Secure, reliable, on-demand access to data, software, people, and other resources (ideally all via a Web Browser)

How it Really Happens (A Simplified View)

Web Browser

Web Portal

Simulation Tool

Data Viewer Tool

Chat Tool

Credential Repository

Certificate authority

Registration Service

Telepresence Monitor

Data Catalog

Compute Server

Compute Server

Camera

Camera

Database service

Database service

Database service

Users *work with client applications*

Application services *organize VOs & enable access to other services*

Collective services *aggregate &/or virtualize resources*

Resources *implement standard access & management interfaces*

How it Really Happens (**with** Grid Software)

Web Browser

CHEF

Simulation Tool

Data Viewer Tool

CHEF Chat Teamlet

MyProxy

Certificate Authority

Globus Index Service

Telepresence Monitor

Globus MCS/RLS

Globus GRAM Compute Server

Globus GRAM Compute Server

Camera

Camera

Globus DAI Database service

Globus DAI Database service

Globus DAI Database service

Application Developer	2
Off the Shelf	9
Globus Toolkit	5
Grid Community	3

Users *work with client applications*

Application services *organize VOs & enable access to other services*

Collective services *aggregate &/or virtualize resources*

Resources *implement standard access & management interfaces*

closer to home…

Increasingly part of the computing for Fermilab experiments is provided off-site

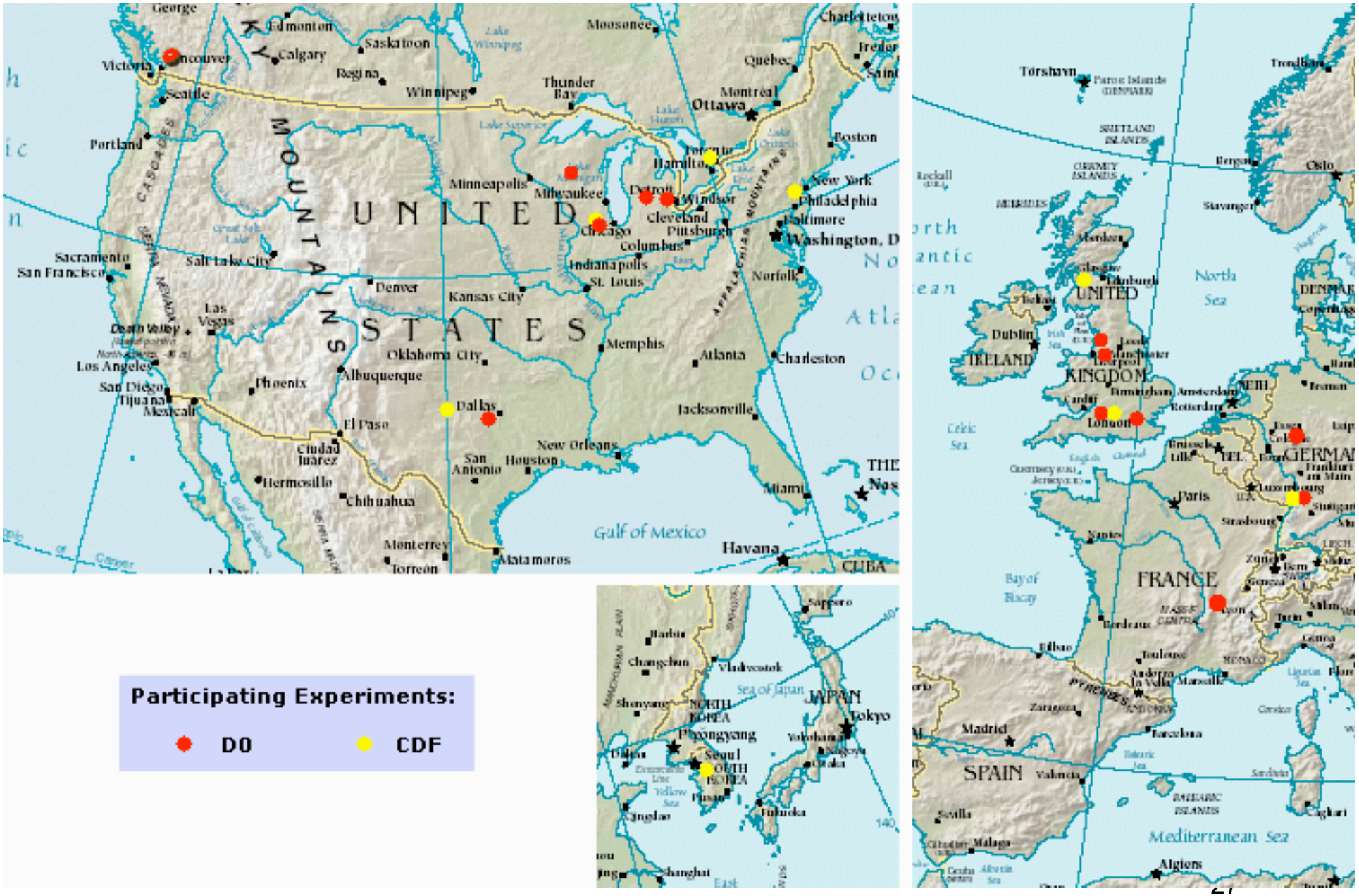
Fermilab has been in the lead in Grid Computing (even before it became a household word)

SAM-GRID is fully functional distributed computing infrastructure in use by D0, CDF and MINOS

- ~30 SAM stations worldwide active for D0
- ~20 SAM stations worldwide active for CDF

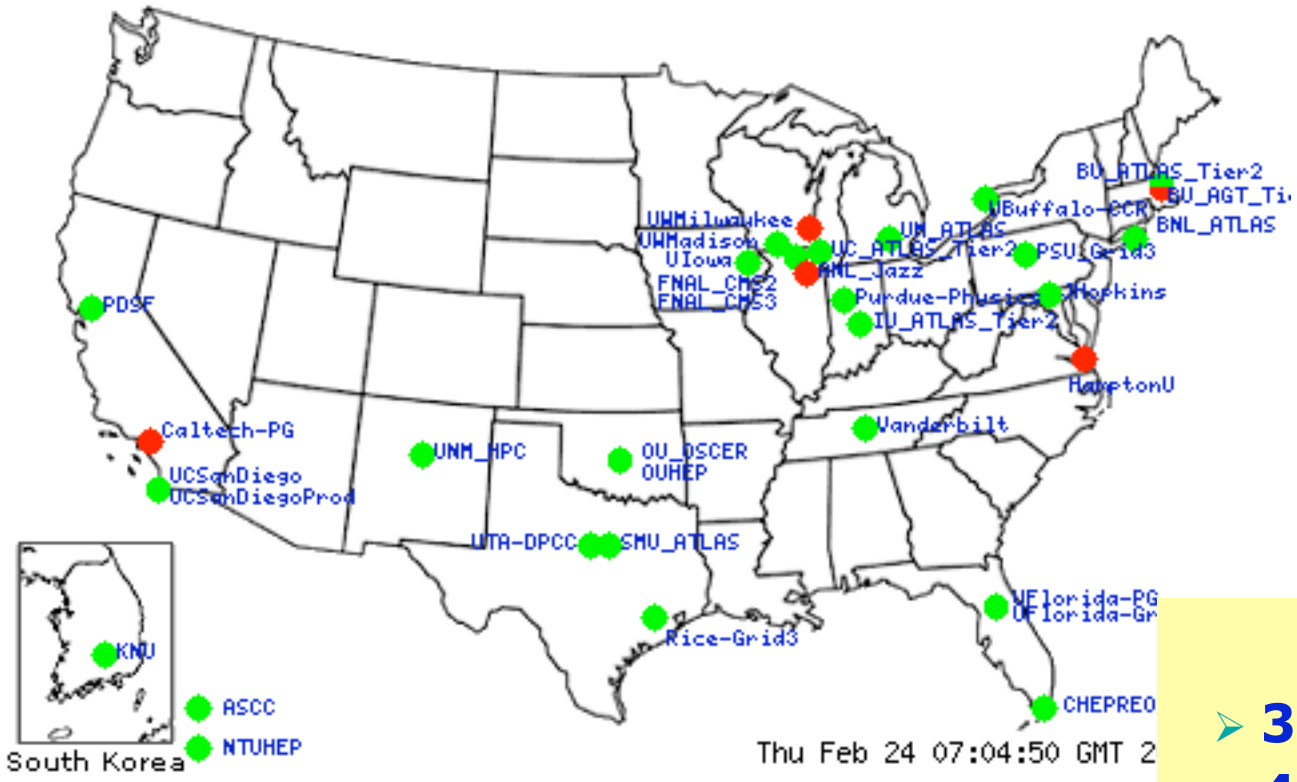
D0 successfully carried out reprocessing of data at 6 sites worldwide

And In order to better serve the entire program of the laboratory the Computing Division will place all of its production computing and storage resources in a Grid infrastructure called FermiGrid.


Participating Experiments:
D0
CDF
UNITED STATES
Gulf of Mexico
Mediterranean Sea
North Sea
Sea of Japan
Bay of Biscay
Celtic Sea
Vancouver
Victoria
Seattle
Portland
Sacramento
San Francisco
Los Angeles
San Diego
Las Vegas
Phoenix
Albuquerque
El Paso
Salt Lake City
Denver
Oklahoma City
Dallas
Houston
San Antonio
Kansas City
Minneapolis
Chicago
Milwaukee
Detroit
Cleveland
Pittsburgh
Columbus
Indianapolis
St. Louis
Memphis
Atlanta
Charleston
Jacksonville
Miami
New Orleans
Washington, D
Philadelphia
New York
Boston
Montreal
Ottawa
Toronto
Hamilton
Windsor
Thunder Bay
Winnipeg
Regina
Saskatoon
Calgary
Edmonton
Moosonee
Québec
Monterrey
Torreón
Matamoros
Hermosillo
Chihuahua
Ciudad Juárez
Havana
CUBA
Tórshavn
Oslo
Stavanger
Bergen
Trondheim
Aberdeen
Glasgow
Edinburgh
Dublin
IRELAND
KINGDOM
London
Amsterdam
Rotterdam
Brussels
Paris
Strasbourg
Luxembourg
Zurich
Geneva
Lyon
Bordeaux
Toulouse
Marseille
Monaco
Bilbao
Madrid
Zaragoza
Barcelona
SPAIN
Valencia
Sevilla
Malaga
Algiers
Sapporo
Vladivostok
Changchun
Shenyang
Pyongyang
Seoul
Busan
Tokyo
Yokohama
Osaka
Fukuoka
Qingdao
Shanghai

Proto-Grid for US LHC

Multi-organization common shared Grid environment



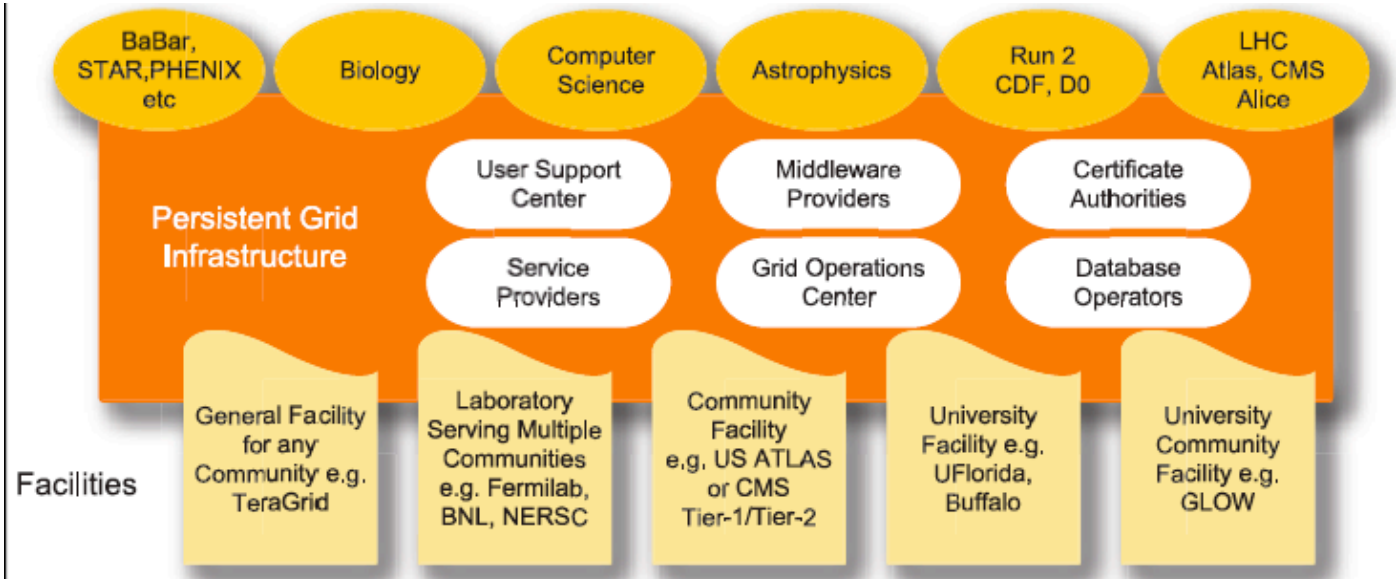
- **35 sites**
- **400-1100 concurrent jobs**
- **10 applications**
- **Running since October 2003**

Plans for Production National Grid Infrastructure
Open Science Grid (OSG)

Join all of the LHC computing resources at labs and universities in the U.S.

Add, over time, computing resources of other high energy and nuclear physics experiments and other scientific partners from Grid Projects

Federate all of these computers and storage systems and services together into a Grid that serves the needs of all of these physics and related disciplines.



Bioinformatics embracing "grid" concepts to change culture of the field

BIOGRID

Biogrid is conducting a trial for the introduction of the GRID approach in the biotechnology industry. The project is focusing on the integration of three existing technologies - agent technology, automatic model classificiation technology and knowledge visualisation technology - and the production of a working prototype biotechnology information GRID.

This Biotechnology information grid will change the perspective of biologists from a single, partial view of biological data towards a holistic view based on data found in document seamlessly integrated with expression and interaction data to model the whole biological network. This constitutes the basis of a next generation research infrastructure for large proteomics and genomics databases.

Contact: ms@biotec.tu-dresden.de
BIOGRID: http://www.bio-grid.net

Workflow - Unified treatment of

What I want to do - likely to evolve in response to new knowledge

What I am doing now - may evolve, e.g., in response to failure

What I did - static, persistent; a source of information

Semantic Grid - Unified treatment of

Describing data - likely to evolve in response to new knowledge

Managing data - may evolve, e.g., in response to failure

Tracking data - static, persistent; a source of information

THE UNIVERSITY of MANCHESTER

Example…

Exploring Williams-Beuren Syndrome using myGrid

Hannah Tipney

Academic Department of Medical Genetics. University of Manchester. UK.

- Contiguous sporadic gene deletion disorder
- 1/20,000 live births, caused by unequal crossover (homologous recombination) during meiosis
- Haploinsufficiency of the region results in the phenotype
- Multisystem phenotype – muscular, nervous, circulatory systems
- Characteristic facial features
- Unique cognitive profile
- Mental retardation (IQ 40-100, mean~60, 'normal' mean ~ 100)
- Outgoing personality, friendly nature, 'charming'

THE UNIVERSITY of MANCHESTER

myGrid

www.mygrid.org.uk

- E-Science pilot research project funded by EPSRC www.mygrig.org.uk
- Manchester, Newcastle, Sheffield, Southampton, Nottingham, EBI and RFCGR, also industrial partners.
- 'targeted to develop open source software to support personalised *in silico* experiments in biology on a grid.'

Which means….

Distributed computing – machines, tools, databanks, people

Personalisation

Provenance and Data management

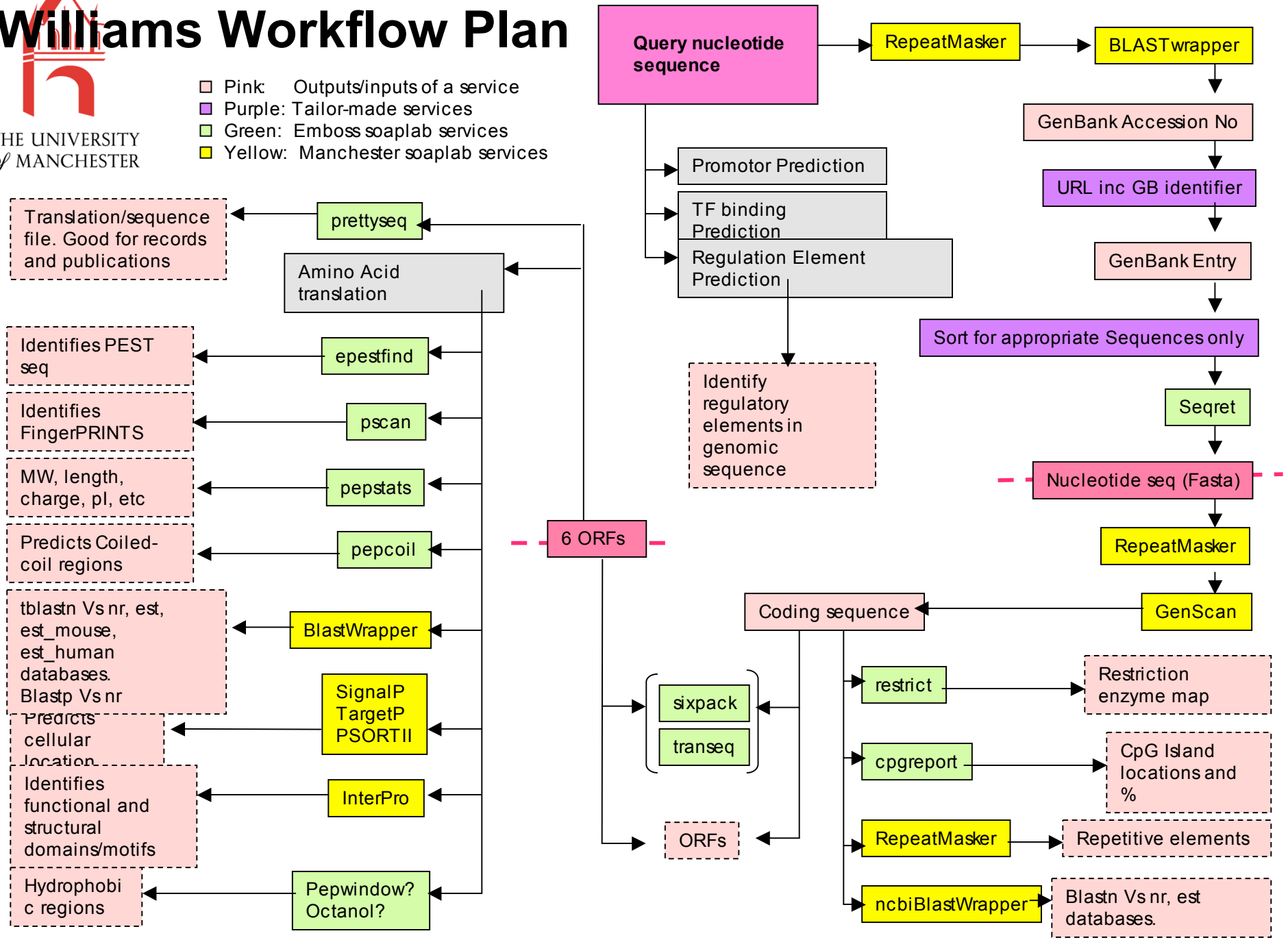
Enactment and notification

A virtual lab 'workbench', a toolkit which serves life science communities.

Williams Workflow Plan

THE UNIVERSITY of MANCHESTER

- Pink: Outputs/inputs of a service
- Purple: Tailor-made services
- Green: Emboss soaplab services
- Yellow: Manchester soaplab services



THE UNIVERSITY of MANCHESTER

The Workflow Experience

Have workflows delivered on their promise? **YES!**

- Correct and Biologically meaningful results
- Automation
  - Saved time, increased productivity
  - Process split into three, you still require humans!
- Sharing
  - Other people have used and want to develop the workflows
- Change of work practises
  - *Post hoc* analysis. Don't analyse data piece by piece receive all data all at once
  - Data stored and collected in a more standardised manner
  - Results amplification
  - Results management and visualisation

The Hype or the Promise

Grid computing appears to be a promising trend for three reasons:

(1) its ability to make more cost-effective use of a given amount of computer resources,

(2) as a way to solve problems that can't be approached without an enormous amount of computing power, and

(3) because it suggests that the resources of many computers can be cooperatively and perhaps synergistically harnessed and managed as a collaboration toward a common objective. ..the computers may collaborate rather than being directed by one managing computer.

One likely area for the use of grid computing will be pervasive computing applications - those in which computers pervade our environment without our necessary awareness.

Search CIO.com Definitions

Democratization of Science

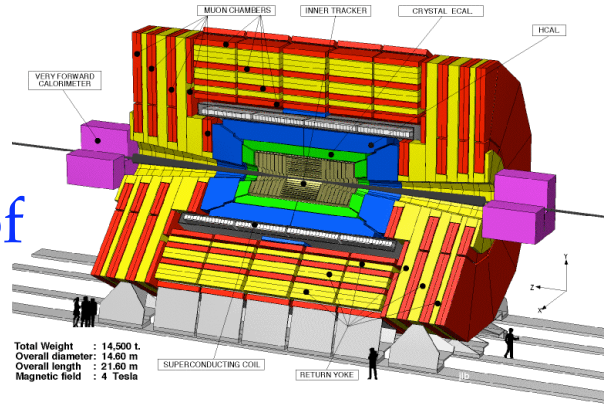
Not just a question of computers and disks and tapes and access to them all

A way of working that puts emphasis on equal access for all and standardization of the way things are done

Enable Scientists to work by creating a massive "virtual" environment (long way to go to get to the vision)

An Evolutionary Change

- Revolution in Science
  - Construct and mine large databases of observational or simulation data
  - Develop simulations & analyses
  - Access specialized devices remotely
  - Exchange information within distributed multidisciplinary teams



- Revolution in Business
  - Enterprise computing is highly distributed, heterogeneous, inter-enterprise (B2B)
  - Business processes increasingly computing- & data-rich
  - Outsourcing becomes feasible => service providers of various sorts

Business Examples 2005

> CSC, Gateway, and IBM are among the companies now using grid technology to provide fee-for-service access to supercomputers.

Walmart inventory control

Satellite technology used to track every item. Inventory adjusted in real time - *data management, prediction, real-time, wide-area synchronization.*

Computer storage: Sun Microsystems said it will offer remote computing resources to business customers, allowing them to purchase computer time over a network as easily as they buy electricity and water through wires and pipes.

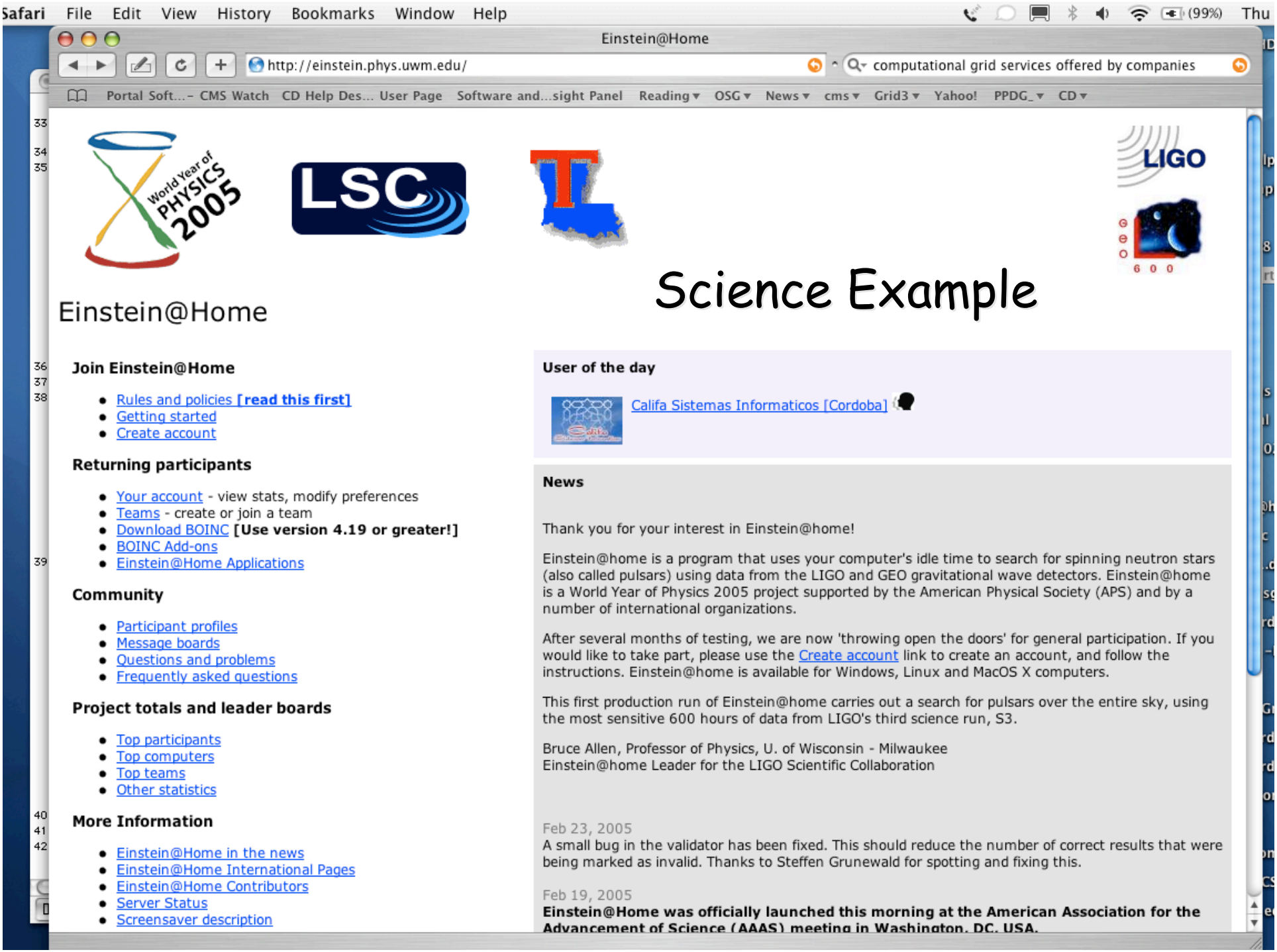
The Sun Grid will cost clients $1 an hour for each microprocessor used and $1 per month for each gigabyte of storage. Customers will pay only for what they use when they use it, Sun said.

Sun (SUNW) said its grid now has 10,000 microprocessors powered up in data centers in Texas, New Jersey, Virginia and Scotland, with more coming online later this year. The server and software company is currently working with pilot customers in both the financial and oil industries, and it plans to make the service more widely available this spring.

By making a large number of computers available to solve a problem and distributing the workload, it can cut a task that once took months to a matter of days. But it's not used in critical jobs such as transferring money to and from bank customers' accounts. Those are still handled by mainframes owned by the banks.

- - -

Compiled by David Cohn. AP and Reuters contributed to this report.


Safari File Edit View History Bookmarks Window Help
(99%) Thu
Einstein@Home
http://einstein.phys.uwm.edu/
computational grid services offered by companies
Portal Soft...- CMS Watch CD Help Des... User Page Software and...sight Panel Reading OSG News cms Grid3 Yahoo! PPDG_ CD
World Year of PHYSICS 2005
LSC
LIGO
Einstein@Home
Science Example
Join Einstein@Home
Rules and policies [read this first]
Getting started
Create account
Returning participants
Your account - view stats, modify preferences
Teams - create or join a team
Download BOINC [Use version 4.19 or greater!]
BOINC Add-ons
Einstein@Home Applications
Community
Participant profiles
Message boards
Questions and problems
Frequently asked questions
Project totals and leader boards
Top participants
Top computers
Top teams
Other statistics
More Information
Einstein@Home in the news
Einstein@Home International Pages
Einstein@Home Contributors
Server Status
Screensaver description
User of the day
Califa Sistemas Informaticos [Cordoba]
News
Thank you for your interest in Einstein@home!
Einstein@home is a program that uses your computer's idle time to search for spinning neutron stars (also called pulsars) using data from the LIGO and GEO gravitational wave detectors. Einstein@home is a World Year of Physics 2005 project supported by the American Physical Society (APS) and by a number of international organizations.
After several months of testing, we are now 'throwing open the doors' for general participation. If you would like to take part, please use the Create account link to create an account, and follow the instructions. Einstein@home is available for Windows, Linux and MacOS X computers.
This first production run of Einstein@home carries out a search for pulsars over the entire sky, using the most sensitive 600 hours of data from LIGO's third science run, S3.
Bruce Allen, Professor of Physics, U. of Wisconsin - Milwaukee
Einstein@home Leader for the LIGO Scientific Collaboration
Feb 23, 2005
A small bug in the validator has been fixed. This should reduce the number of correct results that were being marked as invalid. Thanks to Steffen Grunewald for spotting and fixing this.
Feb 19, 2005
Einstein@Home was officially launched this morning at the American Association for the Advancement of Science (AAAS) meeting in Washington, DC, USA.

Mix of Jobs running on Grid3 for past year